

PATENT
ATTORNEY DOCKET NO. 2792-5177US

NOTICE OF EXPRESS MAILING

Express Mail Mailing Label Number: EL 740545053US

Date of Deposit with USPS: December 20, 2001

Person making Deposit: Orlena Howell

APPLICATION FOR LETTERS PATENT

for

**METHOD AND SYSTEM FOR IMAGE COMPRESSION USING BLOCK SIZE
HEURISTICS**

INVENTOR:

Jonathan J. Dinerstein

Attorney:
Paul C. Oestreich
Registration No. 44,983
TRASKBRITT, P.C.
P.O. Box 2550
Salt Lake City, Utah 84110
(801) 532-1922

METHOD AND SYSTEM FOR IMAGE COMPRESSION USING BLOCK SIZE HEURISTICS

FIELD OF THE INVENTION

[0001] The present invention relates generally to image compression techniques applicable to motion video. More specifically, the present invention includes a method and system for image compression using block size heuristics to improve speed for motion search.

BACKGROUND OF THE INVENTION

[0002] Digital video products and services such as digital satellite service and video streaming over the Internet are becoming increasingly popular and drawing significant attention in the marketplace. Because of limitations in digital signal storage capacity and in network and broadcast bandwidth transmission limitations, there has been a need for compression of digital video signals for efficient storage and transmission of video images. For this reason, many standards for compression and encoding of digital video signals have been developed. For example, the International Telecommunication Union (ITU) has promulgated the H.261, H.263 and H.26L standards for digital video encoding. Additionally, the International Standards Organization (ISO) has promulgated the Motion Picture Experts Group (MPEG) MPEG-1 and MPEG-2 standards for digital video encoding.

[0003] These standards specify with particularity the form of encoded digital video signals and how such signals are to be decoded for presentation to a viewer. However, significant discretion is allowed for selecting how digital video signals are transformed from uncompressed format to a compressed, or encoded format. For this reason, there are many different digital video signal encoders available today. These various digital video signal encoders may achieve varying degrees of compression.

[0004] It is desirable for a digital video signal encoder to achieve a high degree of compression without significant loss of image quality. Video signal compression is generally achieved by representing identical or similar portions of an image as

infrequently as possible to avoid redundancy. A digital motion video image, which may be referred to as a “video stream”, may be organized hierarchically into groups of pictures which includes one or more frames, each of which may represent a single image of a sequence of images of the video stream. All frames may be compressed by reducing redundancy of image data within a single frame. Motion-compensated frames may be further compressed by reducing redundancy of image data within a sequence of frames.

[0005] Motion video compression may be based on the assumption that little change occurs between frames. This is frequently the case for many video signals. This assumption may be used to improve motion video compression because a significant quantity of picture information may be obtained from the previous frame. In this way, only the portions of the picture that have changed need to be stored or transmitted.

[0006] Each video frame may include a number of macroblocks that define respective portions of the video image of the video frame. The term macroblock refers to a “16x16” pixel region. Other block sizes, *i.e.*, 8x16, 16x8, 8x8, 4x8, 8x4 and 4x4, are derived by subdividing the 16x16 macroblock. A motion vector may be used in mapping blocks from one video frame to corresponding blocks of a temporally displaced video frame. A motion vector maps a spatial displacement within the temporally displaced frame of a relatively closely correlated block of picture elements, or pixels. In frames in which subject matter is moving, motion vectors representing spatial displacement may identify a corresponding block that matches a previous block rather closely.

[0007] This is also true when the video sequence includes a camera pan, *i.e.*, a generally uniform spatial displacement of the entirety of the subject matter of the motion video image. In a camera pan, most of the picture information from the previous frame may still be the same, but it may be at a new location in the current picture frame. It is important to know where objects in the current video frame have moved relative to the previous video frame so that as much information can be carried forward from the previous frame as possible. A search to determine where motion has taken place from a reference frame to a current frame is known as “motion estimation”.

[0008] Motion estimation may be obtained by calculating the similarity between two identically placed regions in the previous and current video frames. To calculate the

difference, the sum of absolute differences (SAD) may be used. The result of the SAD is often called “distortion”, as it measures how different two areas of the previous and current frames are. Distortion may be computed as:

$$distortion = \sum |previous(x, y) - current(x, y)| \quad (1)$$

where, *previous(x,y)* is the location of a previous frame of video and *current(x,y)* is the location of a current frame of video. Rate-distortion means to consider not only the similarity in the picture regions, how large of a vector the motion has, *i.e.*, how far an object has traveled. This vector must be stored, and therefore is a cost that must be considered. For this reason, motion estimation is usually performed by a motion search for many nearby locations (*i.e.*, the motion vector is not too long). The optimal solution is found by comparing the rate-distortions of all possible choices.

[0009] Of course, change in the picture from frame to frame will not only happen because of camera motion. Objects within a video frame can also move, *e.g.*, a stationary camera recording a person who is walking past the frame of view. In cases such as this, it is possible that only small regions of the picture have moved, and other small regions have remained in place. Further, for video content such as sports, it's possible for many small objects to be moving in different directions.

[0010] Motion estimation must be capable of dealing with both coarse-grain motion (large objects moving or camera pan) and fine-grain motion (small objects moving). For this reason, H.26L uses 7 different sizes of regions to estimate motion. These are usually called blocks. These sizes include: 16x16, 8x16, 16x8, 8x8, 4x8, 8x4 and 4x4. The larger block sizes are for coarse-grain motion, the smaller block sizes for fine-grain motion. These sizes are in terms of pixels (individual color dots in the picture). However, performing a motion search for all of these block sizes is very expensive. H.26L states that a motion search should be performed for all of them, but we have discovered a better way.

[0011] It is important to note that smaller block sizes are more expensive to store than larger block sizes because each block has a motion vector. In other words, an

entire 16x16 region can be described with a single motion vector, whereas the same region divided into 4x4 blocks needs 16 motion vectors. Because of this and the fact that most motion in video is coarse-grain, the 16x16 block size is usually selected as the best or preferred block size.

[0012] While there are sophisticated methods for performing image compression, they tend to be expensive. Thus, there still exists a need in the art for a method and system for image compression that reduces computational complexity and increases speed of motion video image compression.

SUMMARY OF THE INVENTION

[0013] The present invention includes a method and system for image compression using block size heuristics. A method for motion searching a video frame is disclosed including iteratively decreasing block size until a rate-distortion (*RD*) has been minimized. A method for compressing motion video images is disclosed. Additionally, a system for transmitting and receiving video images is disclosed. The system may be a video conferencing system.

[0014] These embodiments of the present invention will be readily understood by one of ordinary skill in the art by reading the following detailed description in conjunction with the accompanying figures of the drawings.

DESCRIPTION OF THE DRAWINGS

[0015] The drawings illustrate what is currently regarded as a best mode for carrying out the invention. Additionally, like reference numerals refer to like parts in different views or embodiments of the drawings.

[0016] FIG. 1 is a block diagram of a method of compressing a video image in accordance with the present invention.

[0017] FIG. 2A and 2B are a flow chart of a method for motion searching a video frame in accordance with the present invention.

[0018] FIG. 3 is a block diagram of a system for compressing and decompressing images in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0019] The present invention includes a method and system for image compression using block size heuristics. In the following detailed description, for purposes of explanation, specific details are set forth in order to provide a thorough understanding of the present invention. It will be evident, however, to one of ordinary skill in the art that the present invention may be practiced without these specific details.

[0020] FIG. 1 is a block diagram of a method 100 of compressing a video image in accordance with the present invention. Method 100 includes inputting 102 a motion video frame for processing and performing 104 a motion search as discussed in greater detail with regard to FIGS. 2A and 2B, below. Method 100 may also include storing 106 the motion vector for each block in the video frame and residual coding 108 of motion compensated errors. Method 100 may be repeated 110 as shown in FIG. 1 if there are additional frames to process.

[0021] An important aspect of the inventive block size heuristics is that distortion of the video image will increase as the block size increases for a given granularity of motion in a given video image. Conversely, as smaller block sizes are used, the macroblock overhead (motion vectors) will become increasingly and unnecessarily expensive. Therefore, the measure of rate-distortion will usually have a clear minimum for a given granularity of motion in a given video image. In the case of coarse-grain motion, which is most common, the minimum on a rate-distortion curve may be for the 16x16 block size and thus, decreasing block size will only increase rate-distortion. However, if fine-grain motion is taking place, the minimum rate-distortion may be for a block size smaller than the 16x16 macroblock. So, it is advantageous to iteratively search for the minimum rate-distortion and terminate the search soon as the rate-distortion curve begins to increase.

[0022] As previously discussed, different block sizes may be used to compensate for fine-grain and coarse-grain motion. It is known that coarse-grain motion compensation (using large block sizes) is most common. Further, the inventor has discovered that, since large blocks are not as efficient for fine-grain motion, the distortion

as measured by the SAD will be larger than if small block sizes are used. FIGS. 2A and 2B are a flow chart of a presently preferred method 200 of motion searching a video frame in accordance with the present invention.

[0023] Like H.26L, motion searching in accordance with the present invention uses seven block sizes, *i.e.*, 16x16, 8x16, 16x8, 8x8, 4x8, 8x4 and 4x4. Each frame may be partitioned into a number of macroblocks of size 16x16. The macroblocks are subdivided into blocks of sizes 8x16, 16x8, 8x8, 4x8, 8x4 and 4x4 and evaluated for granularity in accordance with the inventive block size heuristic as embodied in method 200. Method 200 is repeated for each macroblock in the current frame. Once all macroblocks have been processed, method 200 may be repeated for a new frame, by incrementing the current frame to a previous frame and obtaining a new frame.

[0024] Method 200 includes performing 202 a motion search for each of the three largest block sizes only, *i.e.*, 16x16, 8x16, and 16x8. Method 200 further includes calculating a rate-distortion (*RD*) for each of the block sizes 16x16, 8x16, and 16x8 and determining 204 whether the *RD* is lowest for the 16x16 block size. If the *RD* of the 16x16 block size is lowest, then coarse-grain motion has taken place from the previous to the present video frame. No more motion searching is performed for this particular macroblock because the block size with the lowest *RD* has been found 216.

[0025] In accordance with the present invention, *RD* may be calculated as follows:

$$RD = n(rate) + m(distortion) \quad (2)$$

where *n* and *m* are scalar values used for weighting *rate* and *distortion*. Selection of the scalar values, *n* and *m*, is within the knowledge of one of ordinary skill in the art and, thus, will not be further elaborated. The *rate* is the number of bits of storage required for macroblock overhead, such as motion vectors. In other words, *rate* is a measure of non-pictorial information that must be sent along with the portion of the image that has changed. For example, a macroblock usually has a few pieces of information associated

with it: (1) the macroblock type and (2) motion vectors. This information is extra overhead, above and beyond whatever pictorial information must be stored.

[0026] The idea behind calculating a *RD* is to measure the overall predicted cost of storage when taking both of these factors (*rate* and *distortion*) into account. The inventive block size heuristic is not dependent on the particular measure of *rate* or *distortion* or the *RD* formed by a linear combination of *rate* or *distortion*. A *rate* is a measure of non-pictorial information overhead. A particular measure of *rate* may be defined as a number of bits of storage required for macroblock overhead. Other measures of *rate* may be suitable in accordance with the present invention

[0027] *Distortion* is an approximation of how much pictorial information must be stored. For example, as more of the picture information in the current differs from the previous video frame, more picture information must be stored. The goal of the motion search is to find the motion vectors and block size that minimizes the *RD* for each macroblock as applied to the current video frame. There are many measures of *distortion* known in the art. A preferred measure of *distortion* in accordance with the present invention is a sum of absolute differences as defined in Eq. (1) above. However, any suitable measure of distortion may be used with the inventive block size heuristic of the present invention.

[0028] Referring again to FIG. 2A, if the 8x16 or 16x8 block size has a lower rate-distortion, then, fine-grain motion is taking place 204. However, the level of granularity is still undetermined and further processing must take place. In other words, smaller block sizes must be motion searched.

[0029] Method 200 may then include performing 206 a motion search for the 8x8 block size and calculating a *RD* for the 8x8 block size. If the 8x8 block size has a smaller *RD* than the previous larger block sizes 208, then the search must be continued because the level of granularity is still uncertain. Alternatively, if the *RD* of the 8x8 block size is larger than that of the previous larger block sizes, the block size with lowest *RD* has been found.

[0030] Method 200 may also include performing 210 a motion search for the 4x8 and 8x4 block sizes and calculating corresponding *RDs*. If one of the 4x8 or 8x4

block sizes has a smaller *RD* than a previous larger block size, the granularity remains uncertain and the search continues. Alternatively, if the *RD* of the 4x8 or 8x4 block sizes is larger than that of the previous larger block sizes, the block size with lowest *RD* has been found 216. Method 200 may also include performing 214 a motion search on the 4x4 block size. At this point a *RD* has been calculated for all block sizes and the block size with the lowest *RD* has been found 216.

[0031] In accordance with the present invention, only the solution with the lowest *RD* is kept and used for further processing in accordance with the method 100 of compressing a video image. A benefit of this technique is that, in most cases, the 16x16 block size is optimal. Therefore, only the 16x16, 8x16, and 16x8 block sizes must be searched in most cases *i.e.*, three out of the seven available block sizes. This may provide a major performance boost.

[0032] An alternative method of motion searching in accordance with the present invention may include selecting one of a plurality of available block sizes to obtain a selected block size, performing a motion search using the selected block size and calculating and storing a rate-distortion for the selected block size. The method may further include determining whether a lowest rate-distortion block size has been found, if not, continuing to search by selecting a next smallest block size if one exists and repeating above starting from performing a motion search using said selected block size.

[0033] FIG. 3 is a block diagram of a system 300 for compressing and decompressing images in accordance with the present invention. System 300 may be configured to implement methods 100 or 200 or both. System 300 may be configured for transmitting and receiving video images. System 300 may be a video conferencing system, for example and not by way of limitation, Sorenson Video 3, available from Sorenson Media, 4393 South Riverboat Road, Suite 300, Salt Lake City, Utah 84123. System 300 may be configured for communication over a network (not shown for clarity). System 300 may include a processor 302 configured for processing computer instructions 306 and a memory 304 for storing computer instructions 306.

[0034] Computer instructions 306 may be in the form of a computer program. System 300 may include computer instructions 306 implementing a method for

compressing motion video images. The method may be method 100 as described above. The method may include inputting a video frame, performing a motion search on the video frame, computing the change between the video frame and a previous video frame not taking into account motion and storing a motion vector for each block in the video frame and the computed change.

[0035] Although this invention has been described with reference to particular embodiments, the invention is not limited to these described embodiments. Rather, the invention is limited only by the appended claims, which include within their scope all equivalent devices or methods that operate according to the principles of the invention as described herein.